BAYESIAN ESTIMATION FOR A PARAMETRIC MARKOV RENEWAL MODEL APPLIED TO SEISMIC DATA

ILENIA EPIFANI, LUCIA LADELLI, AND ANTONIO PIEVATOLO

ABSTRACT. This paper presents a complete methodology for the Bayesian inference of a semi-Markov process, from the elicitation of the prior distribution, to the computation of posterior summaries, including a guidance for its JAGS implementation. The inter-occurrence times (conditional on the transition between two given states) are assumed to be Weibull-distributed. We examine the elicitation of the joint prior density of the shape and scale parameters of the Weibull distributions, deriving in a natural way a specific class of priors, along with a method for the determination of hyperparameters based on "historical data" and moment existence conditions. This framework is applied to data of earthquakes of three types of severity (low, medium and high size) occurred in the central Northern Apennines in Italy and collected by the CPTI04 (2004) catalogue.

1. INTRODUCTION

Markov renewal processes or their semi-Markov representation have been considered in the seismological literature since they are models which allow the distribution of the inter-occurrence times between earthquakes to depend on the last and the next earthquake and to be not necessarily exponential. The time predictable and the slip predictable models studied in Grandori Guagenti and Molina (1986), in Grandori Guagenti *et al.* (1988) and in Betrò *et al.* (1989) are special cases of Markov renewal processes. These models are capable of interpreting the predictable behavior of strong earthquakes in some seismogenetic area. In these processes the magnitude is a deterministic function of the inter-occurrence time. A stationary Markov renewal process with Weibull inter-occurrence times has been studied from a classical statistical point of view in Alvarez (2005). The Weibull model allows to

consider monotonic hazard rates; it contains the exponential model as special case which gives a Markov Poisson point process. In that paper the parameters of the model have been fitted to the large earthquakes in North Anatolian Fault Zone through maximum likelihood and the Markov Poisson point process assumption is tested. In order to capture a non monotonic behavior in the hazard, in Garavaglia and Pavani (2009) the model of Alvarez (2005) is modified and a Markov renewal process with inter-occurrence times that are mixtures of an exponential and a Weibull distribution is fitted to Turkish data. In Masala (2012) a parametric semi-Markov model with generalized Weibull distribution for the inter-occurrence times is considered and fitted to Italian earthquakes. This semi-Markov model with generalized Weibull distributed times has been first used in Foucher *et al.* (2009) to study the evolution of HIV infected patients. Votsi *et al.* (2012) consider a semi-Markov model for the seismic hazard assessment in the Northern Aegean sea and they estimate the quantities of interest (semi-Markov kernel, Markov renewal functions, etc.) through a nonparametric method.

While a wide literature concerning classical inference for Markov renewal models for earthquake forecasting exists, to our knowledge in this context a Bayesian approach is limited. Patwardhan *et al.* (1980) consider a semi-Markov model with log-normal distributed discrete inter-occurrence times that apply to the large earthquakes in the circum-Pacific belt. They stress the fact that using Bayesian techniques is relevant when prior knowledge is available and is fruitful even if the sample size is small. Marín *et al.* (2005) have also employed semi-Markov models in the Bayesian framework but applied to a completely different area: sow farm management. They use WinBugs to perform computations (but without giving details) and elicit their prior distributions on parameters from knowledge on farming practices.

From a probabilistic viewpoint, a Bayesian statistical treatment of a semi-Markov process amounts to model the data as a mixture of Semi-Markov processes, where the mixing measure is supported on the parameters, by means of their prior laws. A complete characterization of such a mixture is given in Epifani *et al.* (2002).

In this paper we develop a parametric Bayesian analysis for a Markov renewal process modeling earthquakes in an Italian seismic region. The magnitudes are classified into three categories according to their severity and these categories represent the state visited by the process. Following Alvarez (2005), the inter-occurrence times are assumed to be Weibull random variables. The prior distribution of the parameters of the model is elicited using an "historical dataset". The time of the last occurrence in the historical dataset constitutes the origin for the "current sample", which is formed by the sequences of earthquakes and the corresponding inter-occurrence times collected up to some time T to form the likelihood function. When T does not coincide with an earthquake the last observed inter-occurrence time is censored. The posterior distribution of the parameters is obtained through Gibbs sampling and the following summaries are estimated: the transition probabilities, the shape and scale parameters of the Weibull waiting times for each transition and the so-called cross state-probabilities (CSPs). The transition probabilities indicate whether the strength of the next earthquake is in some way dependent on the strength of the last one; the shape parameters of the waiting time indicate whether the hazard rate between two earthquakes of given magnitude classes is decreasing or increasing; the CSPs indicate what is the probability that the next earthquake occurs at or before a given time and is of a given magnitude, conditionally on the waiting time from last earthquake and on its magnitude.

The paper is organized as follows. In Section 2 we illustrate the dataset. It is one of the sequences analyzed by Rotondi (2010). Section 3 introduces the parametric Markov renewal model. Three magnitude classes are considered and a Weibull distribution is assumed for the inter-occurrence times. Section 4 deals with the elicitation of the prior, which is based on a generalized inverse Gamma distribution, considering also the case of scarce prior information. Section 5 contains the data analysis with the estimation of the above-mentioned summaries. The matter of whether the observed earthquakes reflect a time predictable or a slip predictable model is also discussed. Section 6 is devoted to some concluding remarks. An appendix contains the detailed derivation of the full conditional distributions and the JAGS implementation of the Gibbs sampler.

FIGURE 1. Map of Italy with dots indicating earthquakes with magnitude $M_w \geq 4.5$ belonging to macroregion MR_3 (Rotondi (2010)). Inclusion in the macroregion was based on the association between events and seismogenetic sources; the region contour has only an aesthetic function.

2. A TEST DATASET

We test our method on a sequence of seismic events chosen among those examined by Rotondi (2010), which was given us by the author. The sequence collects events that occurred in a tectonically homogeneous macroregion, identified as MR_3 by Rotondi (2010) and corresponding to the central Northern Apennines in Italy. The subdivision of Italy into eight (tectonically homogeneous) seismic macroregions can be found in the DISS (2007) and the data are collected by the CPTI04 (2004) catalogue. Considering earthquakes with magnitude $M_w \geq 4.5$, the sequence is complete from year 1838: using a lower magnitude would make the completeness of the series questionable, especially in its earlier part. The map of these earthquakes marked by dots appears in Figure 1. As a lower threshold for the class of strong earhquakes we choose $M_w \geq 5.3$, as suggested by

Rotondi (2010). Then a magnitude state space with three states is obtained by indexing an earthquake by 1, 2 or 3 if its magnitude belongs to intervals $[4.5, 4.9)$, $[4.9, 5.3)$, $[5.3, +\infty)$, respectively. Magnitude 4.9 is just the midpoint between 4.5 and 5.3. Rotondi (2010) considers a nonparametric Bayesian model for the distribution of the inter-occurrence time between strong earthquakes (i.e. $M_w \geq 5.3$), after a preliminary data analysis which rules out Weibull, Gamma, log-normal distributions among other frequently used holding time distributions. On the other hand, with a Markov renewal model, the sequence of all the inter-occurrence times is subdivided into shorter ones according to the magnitudes, so that we think that a parametric distribution is a viable option. In particular, we focussed on the macroregion MR_3 because the Weibull distribution seems to fit the inter-occurrence times better than in other macroregions. This fact is based on qq-plots. The qq-plots for MR_3 are shown in Figure 2. The plot for transitions from 1 to 3 shows a sample quantile that is considerably larger than expected. The outlying point corresponds to a long inter-occurrence time of about 9 years, between 1987 and 1996, while 99 percent of inter-occurence times are below 5 years. Obviously, the classification into macroregions influences the way the earthquake sequence is subdivided.

3. Markov Renewal Model

Let us observe over a period of time $[0, T]$ a process in which different events occur, with random inter-occurence times. Let us suppose that the possible states of the process are the points of the finite set $E = \{1, \ldots, s\}$ and suppose that the process starts in state j_0. Let us denote by τ the number of times the process changes state in the time interval $[0, T]$ and let s_i denote the time of the i-th change of state. Hence, $0 < s_1 < \cdots < s_\tau \leq T$. Let $j_0, j_1, \ldots, j_\tau$ be the sequence of states visited by the process and x_i the sojourn time in state j_{i-1}, for $i = 1, \ldots \tau$. Then

$$(1) \qquad\qquad x_i = s_i - s_{i-1} \qquad \text{for } i = 1, \ldots, \tau$$
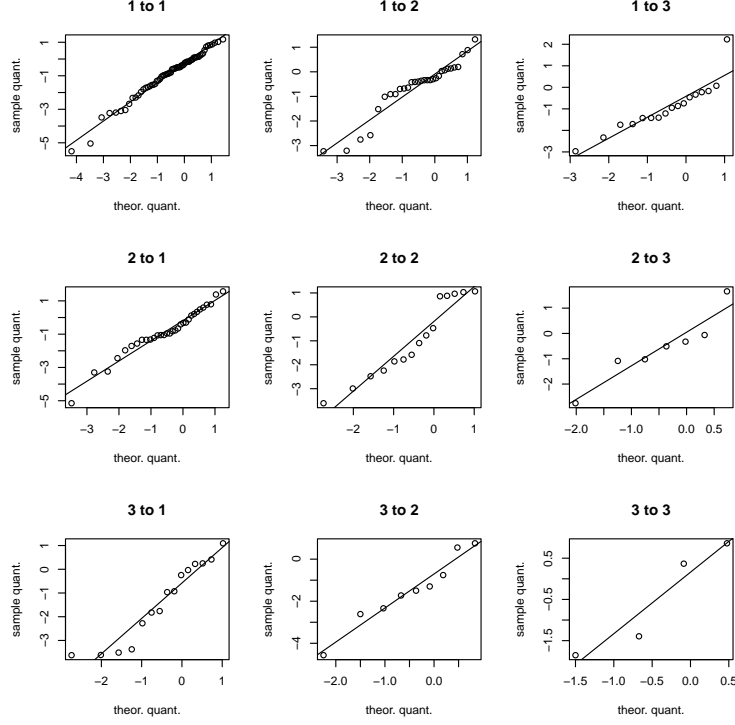
FIGURE 2. Weibull qq-plots of earthquake inter-occurrence times (central Northern Appennines) classified by transition between magnitude classes.

with $s_0 := 0$. Furthermore, let u_T be the time passed in j_τ i.e.

$$(2) \qquad\qquad u_T = T - s_\tau,$$

Time u_T is a right-censored time. Finally, our data are collected in the vector $(\mathbf{j}, \mathbf{x}, u_T)$, where $(\mathbf{j}, \mathbf{x}) = (j_n, x_n)_{n=1,\dots,\tau}$.

In what follows, we assume that the data are the result of the observation, on the time interval $[0, T]$, of an homogeneous Markov renewal process $(J_n, X_n)_{n \geq 0}$ starting from j_0. This means that the sequence $(J_n, X_n)_{n \geq 0}$ satisfies

$$(3) \qquad\qquad P(J_0 = j_0) = 1, \quad P(X_0 = 0) = 1$$

and for every $n \geq 0$, $j \in E$ and $t \geq 0$:

$$(4) \quad P(J_{n+1} = j, X_{n+1} \leq t | (J_k, X_k)_{k \leq n}) = P(J_{n+1} = j, X_{n+1} \leq t | (J_n, X_n)) = p_{J_n j} F_{J_n j}(t),$$

6

where $\mathbf{p} = (p_{ij})_{i,j \in E}$ is a transition matrix and $(F_{ij})_{i,j \in E}$ is an array of distribution functions on $\mathbb{R}_+ = (0, +\infty)$. For more details on Markov Renewal processes see, for example, Limnios and Oprisan (2001). We just recall that, under Assumptions (3) and (4):

- the process $(J_n)_{n \geq 0}$ is a Markov chain, starting from j_0, with transition matrix $\mathbf{p}$,
- the sojourn times $(X_n)_{n \geq 0}$, conditionally on $(J_n)_{n \geq 0}$, form a sequence of independent positive random variables, with distribution function $F_{J_{n-1} \ J_n}$.

We assume the functions F_{ij} are absolutely continuous with respect to the Lebesgue measure with density f_{ij}. Hence, the likelihood function of data $(\mathbf{j}, \mathbf{x}, u_T)$ is:

$$(5) \qquad L(\mathbf{j}, \mathbf{x}, u_T) = \left(\prod_{i=0}^{\tau-1} p_{j_i j_{i+1}} f_{j_i j_{i+1}}(x_{i+1}) \right)^{\mathbb{1}(\tau > 0)} \times \sum_{k \in E} p_{j_\tau k} \bar{F}_{j_\tau k}(u_T),$$

where, for every x, $\bar{F}_{ij}$ is the survival function:

$$\bar{F}_{ij}(x) = 1 - F_{ij}(x) = P(X_{n+1} > x | J_n = i, J_{n+1} = j) \ .$$

Furthermore, we assume that each inter-occurrence time density f_{ij} is a Weibull density with shape parameter α_{ij} and scale parameter θ_{ij}:

$$(6) \qquad f_{ij}(x) = \frac{\alpha_{ij}}{\theta_{ij}} \left(\frac{x}{\theta_{ij}} \right)^{\alpha_{ij}-1} e^{-\left(\frac{x}{\theta_{ij}} \right)^{\alpha_{ij}}} , \quad x > 0, \ \alpha_{ij} > 0, \ \theta_{ij} > 0 \ .$$

For conciseness, let $\boldsymbol{\alpha} = (\alpha_{ij})_{i,j \in E}$ and $\boldsymbol{\theta} = (\theta_{ij})_{i,j \in E}$.

In order to write the likelihood in a more convenient way, let us introduce the following natural statistics. We will say that the process visits the string (i, j) if a visit to state i is followed by a visit to state j and denote:

- x_{ij}^ρ the time spent in state i at the ρ-th visit to string (i, j)
- N_{ij} the number of visits to the string (i, j).

Then, assuming $\tau > 0$, from Equations (5) and (6) we obtain the following expression for the likelihood function:

7

$$(7) \quad L(\mathbf{j}, \mathbf{x}, u_T \,|\, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = \prod_{i,k \in E} p_{ik}^{N_{ik}} \times$$

$$\times \prod_{i,k \in E} \left[\alpha_{ik}^{N_{ik}} \frac{1}{\theta_{ik}^{\alpha_{ik} N_{ik}}} \left(\prod_{\rho=1}^{N_{ik}} x_{ik}^{\rho} \right)^{\alpha_{ik}-1} \times \exp\left\{ -\frac{1}{\theta_{ik}^{\alpha_{ik}}} \sum_{\rho=1}^{N_{ik}} (x_{ik}^{\rho})^{\alpha_{ik}} \right\} \right] \times$$

$$\times \left(\sum_{k \in E} p_{j_\tau k} \exp\left\{ -\left(\frac{u_T}{\theta_{j_\tau k}} \right)^{\alpha_{j_\tau k}} \right\} \right) .$$

Our purpose is now to perform a Bayesian analysis for $\mathbf{p}, \boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ which allows us to introduce prior knowledge on the parameters. As we will show, this analysis is possible, using a Gibbs sampling approach, via the introduction of an auxiliary variable in order to treat easily the right-censored datum.

4. Bayesian analysis

4.1. **The prior distribution.** In our Bayesian analysis, we assume that $\mathbf{p}$ is independent on $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$. In particular, the rows of $\mathbf{p}$ are s independent vectors with Dirichlet distribution with parameters $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_s$ and total mass $c_1, \cdots c_s$, respectively. This means that, for $i = 1, \ldots, s$, the prior density of the i-th row is

$$\pi_{1;i}(p_{i1}, \ldots, p_{is}) = \frac{\Gamma(c_i)}{\prod_{j=1}^{s} \Gamma(\gamma_{ij})} \prod_{j=1}^{s} p_{ij}^{\gamma_{ij}-1}$$

on $T = \{(p_{i1}, \ldots, p_{is}) \,|\, p_{ij} \geq 0, \ \sum_j p_{ij} = 1\}$ where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \cdots, \gamma_{is})$ and $c_i = \sum_{j=1}^{s} \gamma_{ij}$.

As far as $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ is concerned, we will assume the θ_{ij}'s, given the α_{ij}'s, are independent with generalized inverse Gamma density:

$$\pi_{2;ij}(\theta_{ij}|\boldsymbol{\alpha}) = \pi_{2;ij}(\theta_{ij}|\alpha_{ij}) = \frac{\alpha_{ij} b_{ij}(\alpha_{ij})^{m_{ij}}}{\Gamma(m_{ij})} \theta_{ij}^{-(1+m_{ij}\alpha_{ij})} \times \exp\left\{ -\frac{b_{ij}(\alpha_{ij})}{\theta_{ij}^{\alpha_{ij}}} \right\} \qquad \theta_{ij} > 0,$$

where $m_{ij} > 0$ and $b_{ij}(\alpha_{ij})$ is a positive function of α_{ij}. In other terms, a priori $\theta_{ij}^{-\alpha_{ij}}$, given α_{ij}, has a Gamma density with shape m_{ij} and rate $1/b_{ij}(\alpha_{ij}) > 0$. In symbols $\theta_{ij}|\alpha_{ij} \sim \mathcal{GIG}(m_{ij}, b_{ij}(\alpha_{ij}), \alpha_{ij})$.

Finally, we will assume the components of $\boldsymbol{\alpha}$ are independent and have a log-concave density $\pi_{3;ij}$ with support bounded away from zero. As discussed in Gilks and Wild (1992), the log-concavity of $\pi_{3;i,j}$ is necessary in the implementation of the Gibbs sampler (see also Berger and Sun (1993)), although adjustments exist for the non-log-concave case (see Gilks *et al.* (1995)). Conversely, a support bounded away from zero ensures the existence of the posterior moments of the θ_{ij}'s. Anyway, we will specify $\pi_{3;i,j}$ in detail in the next section.

4.2. **Elicitation of the hyperparameters.** In this section we focus our attention on the prior of $(\theta_{ij}, \alpha_{ij})$, for i, j fixed. Following an approach developed in Bousquet (2006), we provide a marginal prior distribution $\pi_{3;i,j}$ for the shapes α_{ij} along with a statistical justification of the choice of the generalized inverse Gamma prior. An interpretation of the hyperparameters is also provided.

First of all, for the sake of clarity, let us drop the indices i, j in the notations and quantities introduced in the previous section. Suppose now that a vector of historical dataset $\mathbf{x}_{-m} = (x_{-m}, \ldots, x_{-1})$ is available, i.e. a past "historical dataset" of m sojourn times in state i followed by a visit to state j. Moreover, let us accept to use as prior density for (α, θ) its posterior density given $\mathbf{x}_{-m}$, when we start from the improper prior:

$$(8) \qquad \tilde{\pi}(\alpha, \theta) \propto \theta^{-1} \left(1 - \frac{\alpha_0}{\alpha}\right)^c \mathbb{1}_{(\theta \geq 0)} \mathbb{1}_{(\alpha \geq \alpha_0)} \ ,$$

with suitable $c \geq 0$ and $\alpha_0 \geq 0$. Then it is immediate to see that this action corresponds to assume as prior distribution for θ given α the density: $\pi_2(\theta|\alpha) = \mathcal{GIG}(m, b(\alpha, \mathbf{x}_{-m}), \alpha)$ and as prior density of α:

$$(9) \qquad \pi_3(\alpha) \propto \frac{\alpha^{m-1-c}(\alpha - \alpha_0)^c}{b^m(\alpha, \mathbf{x}_{-m})} \exp\left(\frac{m\alpha}{\beta(\mathbf{x}_{-m})}\right) \mathbb{1}_{(\alpha \geq \alpha_0)} \ ,$$

with $b(\alpha, \mathbf{x}_{-m}) = \sum_{i=1}^m x_{-i}^\alpha$ and $\beta(\mathbf{x}_{-m}) = m[\sum_{i=1}^m \ln(x_{-i})]^{-1}$.

Remark 1. *We notice that the prior for α, θ we propose has a simple hierarchical structure: $\pi_2(\theta|\alpha)$ is a generalized inverse Gamma density with the first parameter equal to the size of an historical dataset from the same process, whereas the second parameter is a deterministic*

function of α. Moreover, the improper prior (8) we start with depends on two parameters α_0 and c that can be chosen in a such way that the proper prior distribution of θ derived from it has finite moments of sufficiently high order. Finally, it is easy to see that, if $m > 1$, then the prior π_3 of α is proper.

4.3. **Simplifying the prior.** In the previous section we consider a prior with a simple hierarchical structure obtained as the posterior coming from a Bayesian inference on historical data. In order to simplify the expression of the posterior distributions and to ensure that posterior moments are finite, we replace the function $b(\alpha, \mathbf{x}_{-m}) = \sum_{i=1}^{m} x_{-i}^{\alpha}$, in $\pi_2(\theta|\alpha)$ and $\pi_3(\alpha)$ by an easier convex function of α. More precisely, we replace such a function with $b_q(\alpha) = t_q^{\alpha}[(1-q)^{-1/m} - 1]^{-1}$, where t_q is a positive number and $q \in (0,1)$. Then one can see that, for such a choice of $b_q(\alpha)$, t_q turns out to be the quantile of order q of the marginal distribution of the inter-occurrence times between a visit to state i followed by a visit to j. See Bousquet (2010). Hence, for a fixed $q \in (0,1)$, we can estimate t_q using the historical data. Denote by $\hat{t}_q$ such an estimate and let

$$\hat{b}_q(\alpha, \mathbf{x}_{-m}) = \hat{t}_q^{\alpha}[(1-q)^{-1/m} - 1]^{-1}.$$

Thus, we propose for our Bayesian analysis the following priors:

(10) $$\pi_2(\theta|\alpha) = \mathcal{G}IG(m, \hat{b}_q(\alpha, \mathbf{x}_{-m}), \alpha)$$

and

(11) $$\pi_3(\alpha) \propto \alpha^{m-1-c}(\alpha - \alpha_0)^c \exp\left\{-m\alpha\left(\ln \hat{t}_q - \frac{\sum_{i=1}^{m} \ln x_{-i}}{m}\right)\right\} \mathbb{1}_{(\alpha \geq \alpha_0)}.$$

For $m > 1$, $\pi_3(\alpha)$ is a proper prior if q is chosen in such a way that

$$\ln \hat{t}_q - \frac{\sum_{i=1}^{m} \ln x_{-i}}{m} > 0$$

and is always log-concave for any $c \geq 0$. Notice that such a choice is possible for almost all set of data.

10

Finally let us elicit the hyperparameters α_0 and c in order that the posterior second moment of θ is finite. Since

$$\mathbb{E}(\theta^2) = \mathbb{E}\left(\frac{\Gamma(m - 2/\alpha)}{\Gamma(m)}\left[\hat{b}_q(\alpha, \mathbf{x}_{-m})\right]^{2/\alpha}\right) \leq \tilde{K}\mathbb{E}(\Gamma(m - 2/\alpha))$$

for a suitable constant $\tilde{K}$, then one can see that if $\alpha_0 = 2/m$ and $c \geq 1$, then $\mathbb{E}(\theta^2) < +\infty$ and hence too the posterior second moment of θ is finite. As noticed above, $\pi_3(\alpha)$ is a log-concave density for any $c \geq 0$. In particular, for $c = m - 1$, $\pi_3(\alpha)$ is an α_0-shifted Gamma density.

4.4. **Scarce prior information.** The construction of the prior distribution of (α, θ) has to be modified for those pair of states between which no more than two transitions were observed, that is, $m \leq 2$.

If $m = 2$ then $\alpha_0 = 1$, which rules out decreasing hazard rates. In the absence of additional specific prior information, this is an arbitrary restriction, so an α_0 value smaller than 1 must be chosen. Then, the prior second moment of θ is not finite anymore. For the posterior second moment to be finite we need $\alpha > 2/(2 + N)$, where N is the number of transitions between the two concerned states in the (current) sample. Thus the second moment of θ can stay non-finite, even a posteriori, if $2/(2 + N) > \alpha_0$. This would show that the data add little information for that specific transition. To avoid this, we may let $\alpha_0 = 2/3$, corresponding to the smallest prior sample size such that $\alpha_0 < 1$. The value of c can still be $m - 1 = 1$.

If $m = 1$, the single historical observation x_{-1} determines $\hat{b}_q(\alpha, \mathbf{x}_{-m})$. As $\hat{t}_q = x_{-1}$ for any q, it seems reasonable to use $q = 0.5$, so x_{-1} would represent the prior opinion on the median sojourn time. Since with $m = 1$ the argument of the exponent in $\pi_3(\alpha)$ in (11) is zero so that $\pi_3(\alpha)$ is improper, then we make it proper by restricting its support to an interval $[\alpha_0, \alpha_1]$. The value $\alpha_1 = 10$ is suitable for all practical purposes. As before, the choice $\alpha_0 = 2/m = 2$ would be too much restrictive, so we select again $\alpha_0 = 2/3$. The rule $c = m - 1$ would give $c = 0$ in this case, so we keep $c = 1$. If $m = 0$, the prior information on the number of transitions is that there have been no transitions, but there is no information

on the sojourn time. Then we introduce a single random fictitious observation $\tilde{t}_q$, which is uniformly distributed between the smallest and the largest sojourn times associated with all the transitions in the historical dataset, and we use $\tilde{t}_q$ to obtain $\hat{b}_q(\alpha, \tilde{t}_q)$. Thus we fall in the previous case by substituting $m = 1$ for $m = 0$ and by assuming $c = 1$ and $[\alpha_0, \alpha_1] = [2/3, 10]$.

For clearness, we summarize the hyperparameter selection for priors (10)-(11) in Table 1.

	$m > 2$	$m = 2$	$m = 1$	$m = 0$
t_q	$\hat{t}_q$	$\hat{t}_q$	x_{-1}	$\tilde{t}_q$
c	$m - 1$	1	1	1
α_0	$\frac{2}{m}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$

TABLE 1. Hyperparameter selection as the prior sample size m varies

5. ANALYSIS OF THE CENTRAL NORTHERN APENNINES SEQUENCE

In this section we analyze the macroregion MR_3 sequence, using the semi-Markov model. Model fitting, model validation and an attempt at forecasting involves the following steps:

(1) choose the historical dataset $\mathbf{x}_{-m}$ for the elicitation of the prior distribution;
(2) model fit is assessed by comparing observed inter-occurrence times (grouped by transition) to posterior predictive intervals;
(3) cross state-probabilities are estimated, as an indication to the most likely magnitude and time to the next event, given information up to the present time;
(4) an interpretation in terms of slip predictable or time predictable model is provided.

For the elicitation of the prior distribution we cut the observations into two parts, using the first part, called the historical data, for this purpose, and the second part, called the current data, for the likelihood. A look at the matrix of the observed transition frequencies in the whole dataset from year 1838 to 2002, displayed in Table 2(a), shows that our 195-event series contains enough data to allow the assignment of at least three events for any

12

transition to the historical dataset. This choice avoids the need to adjust the prior as summarized in Table 1 for $m_{ij} \leq 2$. The cut-point corresponding to this strategy is the 82nd event which occurred in 1916, with frequency matrix as in Table 2(b).

	(a)				(b)		
	1	2	3		1	2	3
1	65	30	17	1	27	14	4
2	32	15	7	2	14	8	4
3	15	9	4	3	4	3	3

TABLE 2. Number of observed transitions in the whole (a) and historical (b) datasets.

Let us consider the predictive check mentioned above. Fig. 3 shows posterior predictive 95 percent probability intervals of inter-occurrence time for every transition, with the observed inter-occurrence times superimposed. These are empirical intervals computed by generating stochastic inter-occurrence times from their relevant distributions at every iteration of the Gibbs sampler. (The Gibbs sampler itself is completely illustrated in the appendix). Possible outliers, represented as triangles, are those times with Bayesian p-value (that is the predictive tail probability) less than 2.5 percent. The percentage of these extreme points, along with the predictive expected value of the inter-occurrence times, is showed in Table 3. Most points fall within the predictive intervals. Those few inter-occurrence times which are really extreme, such as the small values observed in transitions $(1, 1)$, $(2, 1)$, $(1, 2)$ and the large one in transition $(1, 3)$ match unsurprisingly the outlying points in the corresponding qq-plots in Fig. 2. (In some qq-plots there are more outlying points than in Fig. 3 because the qq-plots comprise the entire series). This fact could be regarded as a lack of fit of the Weibull model, but we are more inclined to attribute it to an imperfect assignment of some events to the macroregion or to an insufficient filtering of secondary events.

The shape parameters α_{ij} are particularly important as they reflect an increasing hazard if larger than 1, a decreasing hazard if smaller than 1 and a constant hazard if equal to 1.

95% posterior predictive intervals of the intertimes
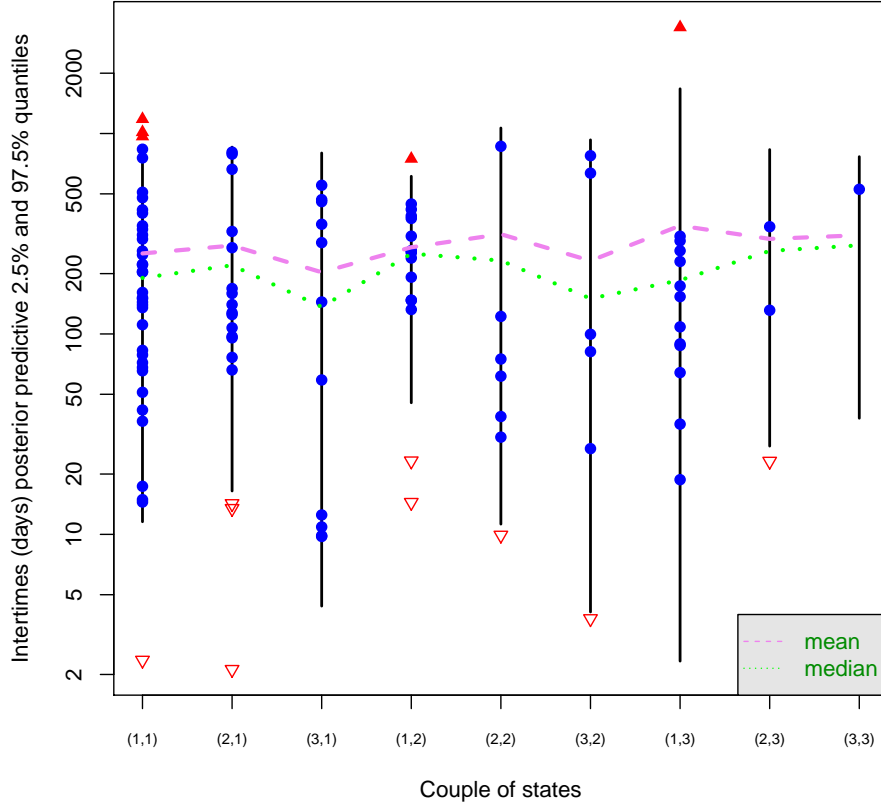
FIGURE 3. Posterior predictive 95 percent credible intervals of the mean inter-occurrence times in days with actual times denoted by (blue) solid dots. Suspect outliers are denoted by (red)-pointing triangles. The (green) dotted line shows the posterior mean and the (violet) dashed line the posterior median.

	1	2	3
1	253.28 (10.5%)	272.27 (18.8%)	340.98 (7.7%)
2	273.45 (16.7%)	314.67 (14.3%)	293.35 (3.3%)
3	204.48 (0.0%)	237.54 (16.7%)	312.73 (0.0%)

TABLE 3. Summaries of the posterior distributions of the mean inter-occurrence times for each transition. Posterior means are followed by (percentage of points having posterior p-value less than 2.5 percent).

	1	2	3
1	1.182 (0.102)	1.981 (0.252)	0.828 (0.103)
2	1.318 (0.173)	1.218 (0.232)	1.716 (0.473)
3	1.026 (0.179)	0.996 (0.150)	2.098 (0.618)

TABLE 4. Summaries of the posterior distributions of the shape parameter α. Posterior means are followed by (standard deviations).

Table 4 displays the posterior means of these parameters (along with their posterior standard deviations). Finally Table 5 shows the posterior means of the transition probabilities.

	1	2	3
1	0.57	0.27	0.16
2	0.58	0.28	0.14
3	0.52	0.32	0.16

TABLE 5. Posterior means of the transition matrix $\mathbf{p}$.

Cross state-probability plots are an attempt at predicting what type of event and when it is most likely to occur. A cross state-probability (CSP) $P^{ij}_{t_0|\Delta x}$ represents the probability that the next event will be in state j within a time interval Δx under the assumption that the previous event was in state i and t_0 time units have passed since its occurrence:

$$(12) \quad P^{ij}_{t_0|\Delta x} = P\left(J_{n+1} = j,\ X_{n+1} \leq t_0 + \Delta x|\ J_n = i,\ X_{n+1} > t_0\right) =$$

$$= \frac{p_{ij}\left(\bar{F}_{ij}(t_0) - \bar{F}_{ij}(t_0 + \Delta x)\right)}{\sum_{h \in E} p_{ih}\bar{F}_{ih}(t_0)} \ .$$

Fig. 4 displays the CSPs with time origin on 31 December 2002, the closing date of the CPTI04 (2004) catalogue. At this time, the last recorded event had been in class 2 and had occurred 965 days earlier (so t_0 is about 32 months). From these plots we can read out the probability that an event of any given type will occur before a certain number of months. For example, after 24 months, the sum of the mean CSPs in the three graphs indicates that the probability that an event will have occurred is more than 90 percent,

with a larger probability assigned to an event of type 2, followed by type 1 and type 3. The posterior means of the CSPs are also reported in Table 6.
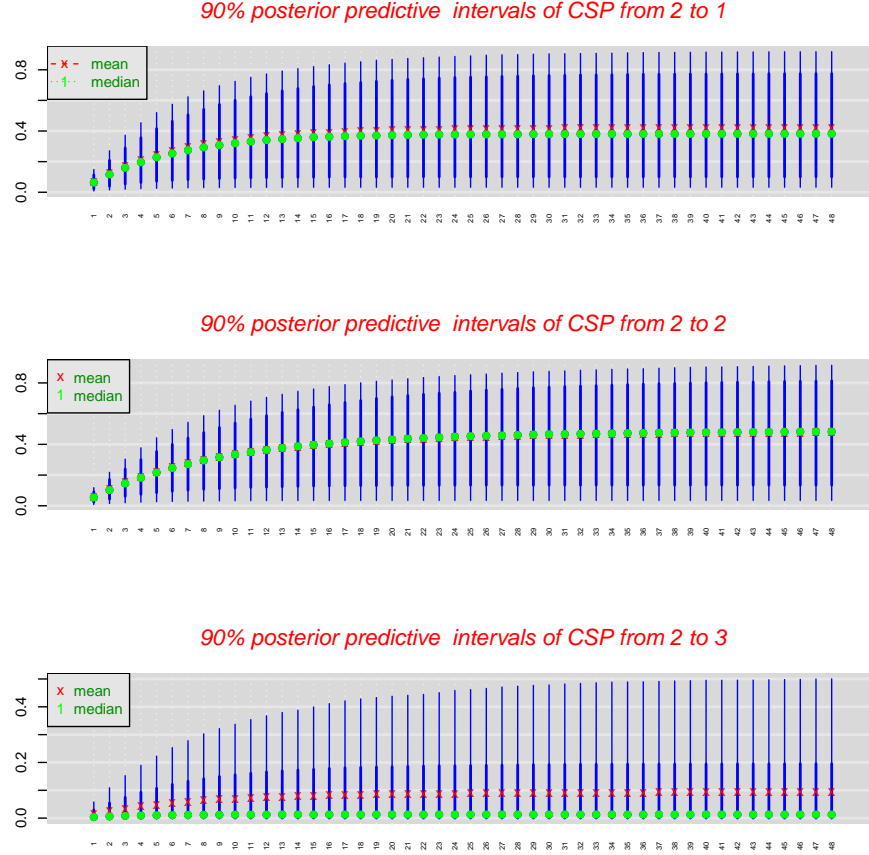
FIGURE 4. Posterior mean and median of CSPs with time origin on 31 December 2002 up to 48 months ahead, along with 90 percent posterior credible intervals. Transitions are from state 2 to state 1, 2 or 3 (first to third panel, respectively).

	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	Year	2 Years	3 Years	4 Years
to 1	0.070	0.122	0.171	0.210	0.243	0.270	0.364	0.410	0.418	0.420
to 2	0.059	0.105	0.150	0.187	0.221	0.249	0.363	0.444	0.468	0.477
to 3	0.014	0.024	0.034	0.041	0.048	0.054	0.075	0.088	0.092	0.093

TABLE 6. CSPs with time origin on 31 December 2002, as represented in Fig. 4.

The predictive capability of our model can be assessed by marking the time of the next event on the relevant CSP plot. In our specific case, the first event in 2003 which can be assigned to the macroregion MR_3 happened in the Forlì area on 26 January and was of type 1, with a CSP of 7 percent. This is a low probability, but a single case is not enough to judge our model, which would be a bad one if repeated comparisons did not reflect the pattern represented by the CSPs. Therefore we repeated the same comparison by re-estimating the model using only the data up to 31 December 2001, 31 December 2000, and so on backwards down to 1992. For this particular exercise the historical data for prior elicitation are such that the frequency of each transition is at least two (not three as done so far) and the hyperparameters are those on the second column of Table 1. The reason is that, by going backwards, the only $(3, 3)$ transition is removed from the current data set, so that the CSP for $(3, 3)$ would rely on prior information only, resulting in a too large posterior variance of the parameters. The results are shown in Table 7. The boxed numbers correspond to the observed events and it is a good sign that they do not always correspond to very high or very low CSPs, as this would indicate that events occur too late or too early compared to the estimated model.

An exception is represented by the last four arrays in the table, which are all related to the period from the last event in July 1987 to the event in October 1996, during which no events with $M_w \geq 4.5$ occurred. This is the longest inter-occurrence time in the whole series and is the outlying point in the top right panel of Fig. 2 and the associated transition is a relatively rare one, so it is not surprising that the boxed CSPs are small.

The examination of the posterior distributions of transition probabilities and of the predictive distributions of the inter-occurrence times can give some insight into the type of energy release and accumulation mechanism. We consider two such mechanisms, the time predictable model (TPM) and the slip predictable model (SPM).

In the TPM, when a maximal energy threshold is reached, some fraction of it (not always the same) is released; therefore, the waiting time for the next event is longer if the current event is stronger. So, the waiting time distribution depends on the current event type, but

end of catalogue: 31/12/2001;   previous event type:2;   waiting time: 600 days

	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	Year	392 days	2 Years	3 Years	4 Years
to 1	0.053	0.095	0.135	0.169	0.199	0.224	0.318	$\boxed{0.373}$	0.375	0.384	0.387
to 2	0.034	0.063	0.092	0.118	0.143	0.166	0.272	0.383	0.388	0.432	0.456
to 3	0.017	0.031	0.043	0.054	0.063	0.070	0.098	0.115	0.115	0.119	0.121

end of catalogue: 31/12/2000;   previous event type: 2;   waiting time: 235 days

	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	Year	2 Years	757 days	3 Years	4 Years
to 1	0.061	0.112	0.162	0.205	0.245	0.280	0.417	0.502	$\boxed{0.505}$	0.518	0.522
to 2	0.022	0.040	0.059	0.076	0.092	0.106	0.176	0.248	0.252	0.279	0.293
to 3	0.017	0.031	0.047	0.061	0.074	0.085	0.132	0.159	0.159	0.164	0.165

end of catalogue: 31/12/1999;   previous event type: 1;   waiting time: 177 days

	1 Month	2 Months	3 Months	4 Months	130 days	5 Months	6 Months	Year	2 Years	3 Years	4 Years
to 1	0.062	0.115	0.166	0.210	0.222	0.251	0.286	0.430	0.525	0.544	0.548
to 2	0.038	0.074	0.110	0.142	$\boxed{0.151}$	0.172	0.197	0.283	0.305	0.305	0.305
to 3	0.013	0.023	0.033	0.042	0.045	0.050	0.058	0.091	0.122	0.135	0.140

end of catalogue: 31/12/1998;   previous event type: 3;   waiting time: 280 days

	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	188 days	Year	2 Years	3 Years	4 Years
to 1	0.062	0.109	0.153	0.189	0.221	0.247	$\boxed{0.252}$	0.339	0.389	0.400	0.402
to 2	0.037	0.067	0.095	0.119	0.141	0.160	0.164	0.234	0.290	0.307	0.313
to 3	0.037	0.068	0.100	0.126	0.150	0.170	0.175	0.238	0.268	0.274	0.276

end of catalogue: 31/12/1997;   previous event type: 3;   waiting time: 442 days

	1 Month	2 Months	85 days	3 Months	4 Months	5 Months	6 Months	Year	2 Years	3 Years	4 Years
to 1	0.074	0.131	0.176	0.184	0.227	0.264	0.294	0.402	0.463	0.475	0.479
to 2	0.054	0.096	0.131	0.138	0.173	0.205	0.232	0.343	0.428	0.455	0.466
to 3	0.010	0.016	$\boxed{0.021}$	0.021	0.025	0.028	0.030	0.037	0.041	0.042	0.043

end of catalogue: 31/12/1996;   previous event type: 3;   waiting time: 77 days

	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	Year	450 days	2 Years	3 Years	4 Years
to 1	0.073	0.131	0.186	0.231	0.271	0.304	0.422	0.447	0.483	0.493	0.496
to 2	0.043	0.075	0.106	0.132	0.155	0.174	0.248	0.268	0.300	0.314	0.319
to 3	0.012	0.027	0.049	0.076	0.105	0.129	0.172	$\boxed{0.175}$	0.178	0.179	0.179

end of catalogue: 31/12/1995;   previous event type: 1;   waiting time: 3100 days

	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	288 days	Year	2 Years	3 Years	4 Years
to 1	0.173	0.304	0.420	0.511	0.589	0.651	0.798	0.861	0.964	0.982	0.985
to 2	0.001	0.002	0.002	0.003	0.003	0.004	0.004	0.004	0.005	0.005	0.005
to 3	0.001	0.003	0.004	0.004	0.005	0.005	$\boxed{0.007}$	0.007	0.008	0.008	0.008

end of catalogue: 31/12/1994;   previous event type: 1;   waiting time: 2735 days

	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	Year	653 days	2 Years	3 Years	4 Years
to 1	0.169	0.295	0.410	0.502	0.580	0.643	0.857	0.954	0.964	0.982	0.986
to 2	0.001	0.002	0.002	0.003	0.003	0.004	0.005	0.005	0.005	0.005	0.005
to 3	0.001	0.002	0.003	0.004	0.004	0.005	0.007	$\boxed{0.007}$	0.008	0.008	0.008

end of catalogue: 31/12/1993;   previous event type: 1;   waiting time: 2370 days

	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	Year	2 Years	1018 days	3 Years	4 Years
to 1	0.166	0.289	0.403	0.495	0.573	0.636	0.854	0.964	0.981	0.982	0.986
to 2	0.001	0.002	0.003	0.003	0.004	0.004	0.005	0.005	0.005	0.005	0.005
to 3	0.001	0.002	0.003	0.003	0.004	0.004	0.006	0.007	$\boxed{0.007}$	0.007	0.007

end of catalogue: 31/12/1992;   previous event type: 1;   waiting time: 2005 days

	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	Year	2 Years	3 Years	1383 days	4 Years
to 1	0.161	0.283	0.395	0.486	0.564	0.628	0.849	0.963	0.983	0.986	0.987
to 2	0.001	0.002	0.003	0.003	0.004	0.004	0.005	0.006	0.006	0.006	0.006
to 3	0.001	0.002	0.002	0.003	0.003	0.004	0.005	0.006	0.006	$\boxed{0.006}$	0.006

TABLE 7. CSPs as the end of the catalogue shifts back by one-year steps. The numbers in boxes are the probability that the next observed event has occurred at or before the time when it occurred and is of the type that has been observed.

not on the next event type, that is, we expect $F_{ij}(t) = F_{i\cdot}(t)$, $j = 1, 2, 3$. The strength of an event, does not depend on the strength of the previous one, because every time the same energy level has to be reached for the event to occur, so we expect $p_{ij} = p_{\cdot j}$, $j = 1, 2, 3$, that is, a transition matrix with equal rows (which is indeed the case, as seen from Table 5). The CSPs (12) would factorize as follows,

$$(13) \qquad P^{ij}_{t_0|\Delta x} = \frac{p_{ij}\left(\bar{F}_{ij}(t_0) - \bar{F}_{ij}(t_0 + \Delta x)\right)}{\sum_{h \in E} p_{ih}\bar{F}_{ih}(t_0)} = \frac{p_{\cdot j}\left(\bar{F}_{i\cdot}(t_0) - \bar{F}_{i\cdot}(t_0 + \Delta x)\right)}{\bar{F}_{i\cdot}(t_0)} ,$$

so that, given i, they are proportional to each other as $j = 1, 2, 3$ for any Δx.

In the SPM, after an event energy falls to a minimal threshold and increases until the next event, where it starts to increase again from the same threshold. Here, the magnitude of an event depends on the length of the waiting time, but not on the magnitude of the previous one, because energy always accumulates from the same threshold. In this case again $p_{ij} = p_{\cdot j}$, but $F_{ij}(t) = F_{\cdot j}(t)$, so

$$(14) \qquad P^{ij}_{t_0|\Delta x} = \frac{p_{\cdot j}\left(\bar{F}_{\cdot j}(t_0) - \bar{F}_{\cdot j}(t_0 + \Delta x)\right)}{\sum_{h \in E} p_{\cdot h}\bar{F}_{\cdot h}(t_0)} .$$

If this is the case CSPs, given j would be equal to each other as $i = 1, 2, 3$ for any Δx.

As the a posteriori transition probability matrix has essentially equal rows, the CSPs should be examined with regard to the properties (13) and (14) to decide whether our sequence is closer to a TPM or to an SPM. An additional feature that can help discriminate is the tail of the waiting time distribution: for a TPM, the tail of the waiting time distribution is thinner *after* a weak earthquake than *after* a strong one; for an SPM, the tail of the waiting time is thinner *before* a weak earthquake than *before* a strong one. However this additional job proves unnecessary, after examination of Figures 5 and 6, which represent the posterior means of the ratios of the CSPs as a function of Δx for $t_0 = 0$. In Fig. 5, the ratio of two CSPs –say $P^{ij}_{t_0|\Delta x}/P^{ik}_{t_0|\Delta x}$– starts to be equal to p_{ij}/p_{ik} only when Δx is large enough for the survival functions $\bar{F}_{ij}(t_0 + \Delta x)$ and $\bar{F}_{ik}(t_0 + \Delta x)$ to be close to zero. A similar behaviour is observed in Fig. 6. The conclusion is that neither a SPM nor a TPM

is supported by the posterior distributions of the parameters, because of the behavior of the waiting times.

6. Concluding Remarks

Whe have presented a complete methodology for the Bayesian inference of a semi-Markov process, from the elicitation of the prior distribution, to the computation of posterior summaries, including a guidance for its JAGS implementation. In particular, we have examined in detail the elicitation of the joint prior density of the shape and scale parameters of the Weibull-distributed holding times (conditional on the transition between two given states), deriving in a natural way a specific class of priors, along with a method for the determination of hyperparameters based on "historical data" and moment existence conditions. This framework has been applied to the analysis of seismic data, but it can be adopted for inference on any system for which a Markov renewal process model is plausible. A possible and not-yet explored application is the modelling of voltage sags (or voltage dips) in power engineering: the state space would be formed by different classes of voltage, starting from voltage around its nominal value, down to progressively deeper sags. In the engineering literature, the dynamic aspect of this problem is in fact disregarded, while it could help bringing additional insight into this phenomenon.

With regard to the seismic data analysis, other uses of our model can be envisaged. The model can be applied to areas with a less complex tectonics, such as Turkey, by replicating for example Alvarez's anaysis. Outliers, such as those appearing in Figure 3, could point out to events whose assignment to a specific seismogenetic source should be re-discussed. The analysis of earthquake occurrence can support decision making related to the risk of future events. We have not examined this issue here, but a methodology is outlined by Cano *et al.* (2011).

A final note concerns the more recent Italian seismic catalogue CTPI11 (2011), including events up to the end of 2006. Every new release of the catalogue involves numerous changes in the parameterization of earthquakes; as the DISS event classification by macroregion is
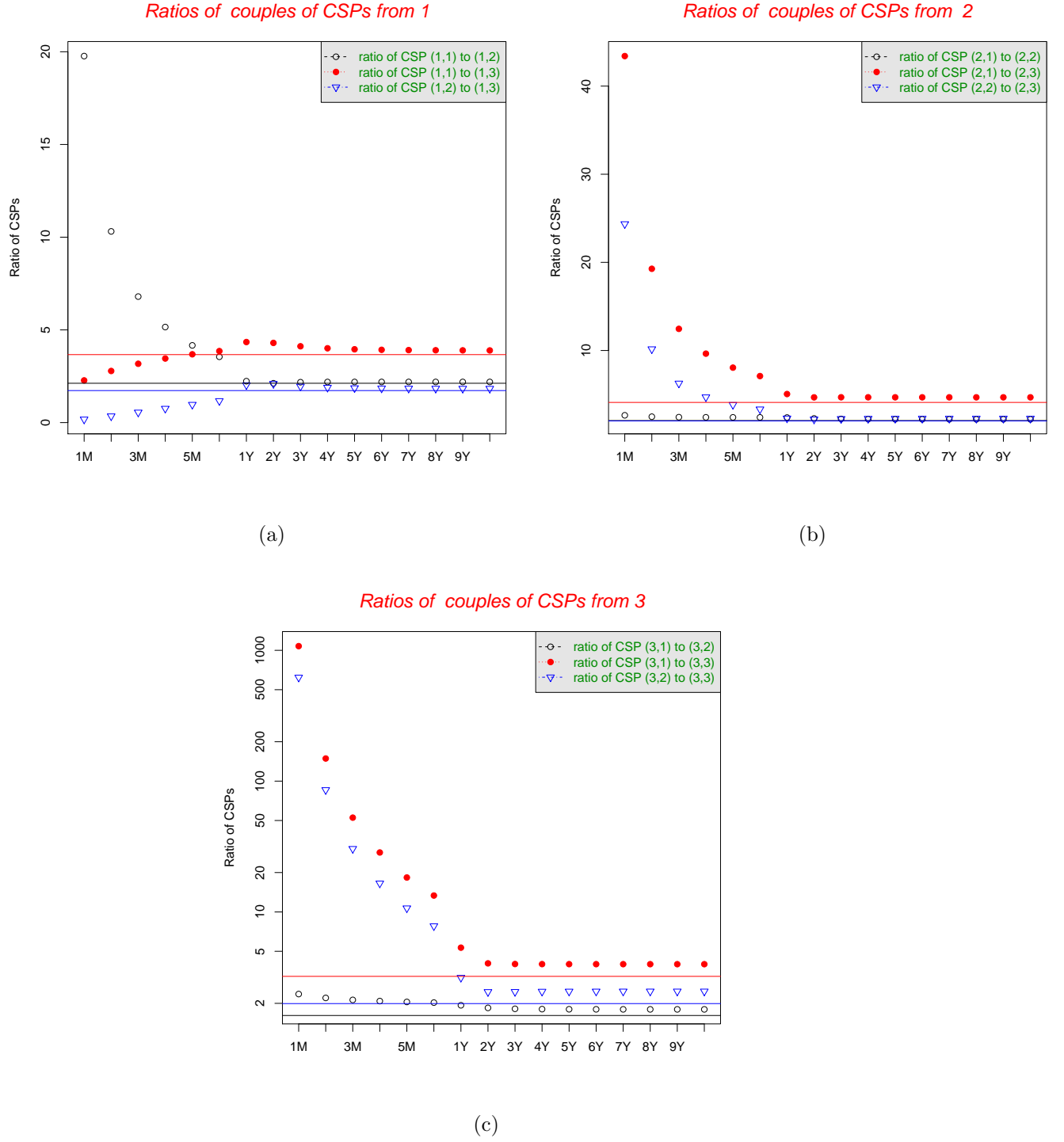
(a)

(b)

(c)

FIGURE 5. Posterior means of ratios of CSPs $P^{ij}_{0|\Delta x}/P^{ik}_{0|\Delta x}$ with time origin on 0 up to 10 years ahead. Transitions are from state 1, 2 and 3 to state 1, 2 or 3.
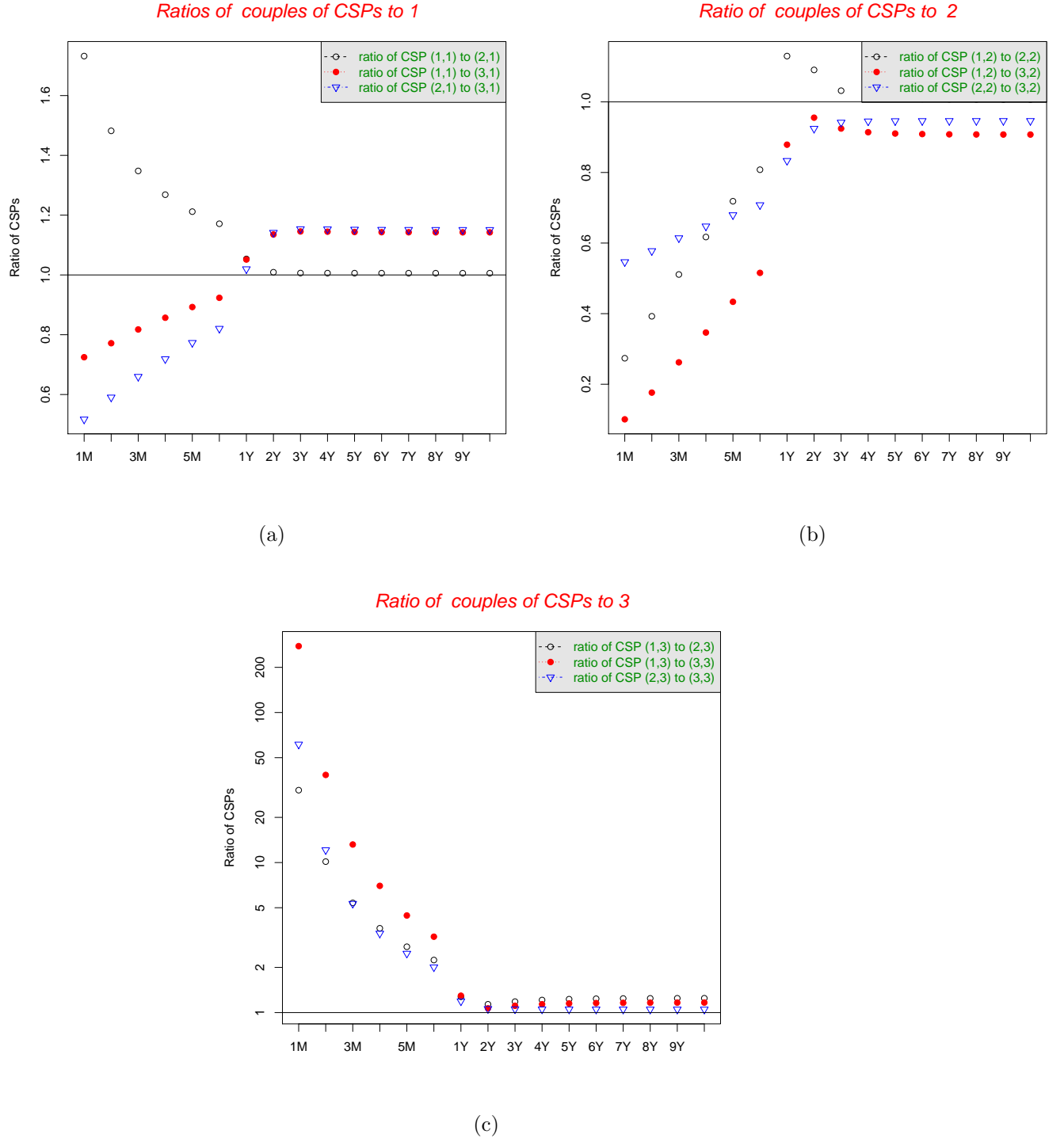
21

(a)

(b)

(c)

FIGURE 6. Posterior means of ratios of CSPs $P^{ij}_{0|\Delta x}/P^{kj}_{0|\Delta x}$ with time origin on 0 up to 10 years ahead. Transitions are from state 1, 2 and 3 to state 1, 2 or 3.

not yet available for events in this catalogue we could not use this more recent source of data.

APPENDIX A. GIBBS SAMPLING

Here we derive the full conditional distributions needed to implement a Gibbs sampling and give indications on its JAGS (Plummer 2010) implementation.

A.1. **Full conditional distributions.** Without loss of generality we can assume that the last holding time is censored i.e. $u_T > 0$. In this case in order to obtain simple full conditional distributions and then an efficient Gibbs sampling, we introduce the auxiliary variable $j_{\tau+1}$ which represents the unobserved state following j_τ. Thus the state space of the Gibbs sampler is $(\mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}, j_{\tau+1})$ and the following full likelihood derived from (7):

$$(15) \quad L(\mathbf{j}, \mathbf{x}, u_T, j_{\tau+1} | \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = \prod_{i,k \in E} p_{ik}^{N_{ik}} \times$$

$$\times \prod_{i,k \in E} \left[\alpha_{ik}^{N_{ik}} \frac{1}{\theta_{ik}^{\alpha_{ik} N_{ik}}} \left(\prod_{\rho=1}^{N_{ik}} x_{ik}^\rho \right)^{\alpha_{ik}-1} \times \exp\left\{ -\frac{1}{\theta_{ik}^{\alpha_{ik}}} \sum_{\rho=1}^{N_{ik}} (x_{ik}^\rho)^{\alpha_{ik}} \right\} \right] \times$$

$$\times \left(p_{j_\tau j_{\tau+1}} \exp\left\{ -\left(\frac{u_T}{\theta_{j_\tau j_{\tau+1}}} \right)^{\alpha_{j_\tau j_{\tau+1}}} \right\} \right)$$

is multiplied by the prior and used to determine the full conditionals. For every i and j let

$$\mathbf{p}_{(-i)} = \text{the transition matrix } \mathbf{p} \text{ without } i\text{-th row,}$$

$$\boldsymbol{\alpha}_{(-ij)} = (\alpha_{hk}, \quad h, k \in E, \quad (h,k) \neq (i,j)),$$

$$\boldsymbol{\theta}_{(-ij)} = (\theta_{hk}, \quad h, k \in E, \quad (h,k) \neq (i,j)),$$

$$\tilde{N}_{ij} = N_{ij} + \mathbb{1}_{((j_\tau, j_{\tau+1})=(i,j))}, \quad \tilde{\mathbf{N}}_i = \left(\tilde{N}_{ij}, \quad j = 1, \dots, s \right),$$

$$\tilde{M}_{ij}(\alpha_{ij}) = \sum_{\rho=1}^{N_{ij}} (x_{ij}^\rho)^{\alpha_{ij}} + u_T^{\alpha_{ij}} \mathbb{1}((j_\tau, j_{\tau+1}) = (i,j)),$$

$$\mathcal{C}_{ij} = \prod_{\rho=1}^{N_{ij}} x_{ij}^\rho .$$

Furthermore, let m_{ij} be the number of visits to the string (i,j) in the historical dataset, $x_{ij}^{-\rho}$, for $\rho = 1,\dots m_{ij}$, denote the corresponding inter-occurrence times and $\mathbf{x}_{-m_{ij}} = (x_{ij}^{-m_{ij}},\dots,x_{ij}^{-1})$.

The following result on the full conditional distributions of the Gibbs sampling holds.

Proposition A.1. *Let the prior on* $(\mathbf{p},\ \boldsymbol{\alpha},\ \boldsymbol{\theta})$ *be as in Section 4. Then*

(a) the conditional distribution of $\mathbf{p}_i$, *given* $\mathbf{j}$, $\mathbf{x}$, u_T, $j_{\tau+1}, \mathbf{p}_{(-i)}$, $\boldsymbol{\alpha}$ *and* $\boldsymbol{\theta}$, *is a Dirichlet distribution with parameter* $\tilde{\mathbf{N}}_i + \boldsymbol{\gamma}_i$;

(b) the conditional distribution of $\theta_{ij}^{\alpha_{ij}}$, *given* $\mathbf{j}$, $\mathbf{x}$, u_T, $j_{\tau+1}$, $\mathbf{p}$, $\boldsymbol{\alpha}$ *and* $\boldsymbol{\theta}_{(-ij)}$ *is an inverse-Gamma distribution of parameters* $m_{ij} + N_{ij}$ *and* $\hat{b}_q(\alpha_{ij},\mathbf{x}_{-m_{ij}}) + \tilde{M}_{ij}(\alpha_{ij})$;

(c) the conditional density of α_{ij}, *given* $\mathbf{j}$, $\mathbf{x}$, u_T, $j_{\tau+1}$, $\mathbf{p}$, $\boldsymbol{\alpha}_{(-ij)}$ *and* $\boldsymbol{\theta}$ *is log-concave and, up to a multiplicative constant, is equal to:*

$$(16) \qquad \pi_{3;ij}(\alpha_{ij}) \times \left(\mathcal{C}_{ij} t_{q_{ij}}^{ij}\ \theta_{ij}^{-m_{ij}-N_{ij}} \right)^{\alpha_{ij}} \alpha_{ij}^{N_{ij}+1} \exp\left\{ -\frac{\hat{b}_q(\alpha_{ij},\mathbf{x}_{-m_{ij}}) + \tilde{M}_{ij}(\alpha_{ij})}{\theta_{ij}^{\alpha_{ij}}} \right\} ,$$

where $t_{q_{ij}}^{ij}$ *is determined from Table 1;*

(d) the conditional density of the unseen state $J_{\tau+1}$, *given* $\mathbf{j}$, $\mathbf{x}$, u_T, $\mathbf{p}$, $\boldsymbol{\alpha}$ *and* $\boldsymbol{\theta}$, *is*

$$\frac{p_{j\tau j} \exp\left\{ -\left(\frac{u_T}{\theta_{j\tau j}}\right)^{\alpha_{j\tau j}} \right\}}{\sum_{k\in E} p_{j\tau k} \exp\left\{ -\left(\frac{u_T}{\theta_{j\tau j}}\right)^{\alpha_{j\tau j}} \right\}} ;$$

(e) if $m_{ij} > 0$ *then* $t_{q_{ij}}^{ij}$ *is a known constant (see Table 1). If* $m_{ij} = 0$, *then the conditional distribution of* $\tilde{t}_{q_{ij}}^{ij}$ *given* $\mathbf{j}$, $\mathbf{x}$, u_T, $\mathbf{p}$, $\boldsymbol{\alpha}$ *and* $\boldsymbol{\theta}$ *is a double-truncated Weibull with shape* $\alpha_{ij} + 1$, *scale* θ_{ij} *and the truncation points given by the smallest and the largest inter-occurrence times included in all the historical dataset.*

Proof. Notice that the row $\mathbf{p}_i$, given the data and $j_{\tau+1}$ is conditionally independent on $(\mathbf{p}_{(-i)},\boldsymbol{\alpha},\boldsymbol{\theta})$. Hence

$$\pi(\mathbf{p}_i|\mathbf{j},\mathbf{x},u_T,\ j_{\tau+1},\mathbf{p}_{(-i)},\boldsymbol{\alpha},\boldsymbol{\theta}) \propto L(\mathbf{j},\mathbf{x},u_T,\ j_{\tau+1}|\mathbf{p},\boldsymbol{\alpha},\boldsymbol{\theta}) \times \pi_{1;i}(\mathbf{p}_i) \propto \prod_{j\in E} p_{ij}^{\tilde{N}_{i,j}} \times \prod_{j\in E} p_{ij}^{\gamma_{ij}-1}$$

and point (a) of the thesis follows.

With regard to the full conditional distribution of θ_{ij} given $(\mathbf{j}, \mathbf{x}, u_T,\ j_{\tau+1}, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}_{(-ij)})$, we have

$$\pi_{2;ij}(\theta_{ij}|\mathbf{j}, \mathbf{x}, u_T,\ j_{\tau+1}, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}_{(-ij)}) \propto L(\mathbf{j}, \mathbf{x}, u_T,\ j_{\tau+1}\ |\mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}) \times \pi_{2;ij}(\theta_{ij}|\alpha_{ij})$$

$$= \prod_{i,k \in E} p_{ij}^{\tilde{N}_{ij}} \times \prod_{i,k \in E} \left[\alpha_{ik}^{N_{ik}} \frac{\mathcal{C}_{ik}^{\alpha_{ik}-1}}{\theta_{ik}^{\alpha_{ik}N_{ik}}} \times \mathrm{e}^{-\frac{\tilde{M}_{ik}(\alpha_{ik})}{\theta_{ik}^{\alpha_{ik}}}} \right] \times \frac{\alpha_{ij}\hat{b}_q(\alpha_{ij}, \mathbf{x}_{-m_{ij}})^{m_{ij}}}{\Gamma(m_{ij})} \theta_{ij}^{-(1+m_{ij}\alpha_{ij})}$$

$$\times \exp\left\{ -\frac{\hat{b}_q(\alpha_{ij}, \mathbf{x}_{-m_{ij}})}{\theta_{ij}^{\alpha_{ij}}} \right\}$$

$$\propto \theta_{ij}^{-[1+\alpha_{ij}(m_{ij}+N_{ij})]} \exp\left\{ -\frac{\hat{b}_q(\alpha_{ij}, \mathbf{x}_{-m_{ij}}) + \tilde{M}_{ij}(\alpha_{ij})}{\theta_{ij}^{\alpha_{ij}}} \right\}.$$

As one can see, the last function is the kernel of an inverse Gamma distribution with parameters $m_{ij} + \theta_{ij}$ and $\hat{b}_q(\alpha_{ij}, \mathbf{x}_{-m_{ij}}) + M_{i,j}(\alpha_{ij})$ and point (b) of the thesis follows.

A similar reasoning yelds a posterior distribution $\pi_{3;ij}(\alpha_{ij}|\mathbf{j}, \mathbf{x}, u_T,\ j_{\tau+1}, \mathbf{p}, \boldsymbol{\alpha}_{(-ij)}, \boldsymbol{\theta})$ which is proportional to (16). Furthermore, concerning its log-concavity, from (16) it follows that

$$\log \pi(\alpha_{ij}|\mathbf{j},\ \mathbf{x},\ u_T,\ j_{\tau+1},\ \mathbf{p},\ \boldsymbol{\alpha}_{(-ij)},\ \boldsymbol{\theta}) =$$

$$= \log \pi_{3;ij}(\alpha_{ij}) + \alpha_{ij} \left[\log(\mathcal{C}_{ij} t_{q_{ij}}^{ij})\ - (m_{ij} + N_{ij}) \log \theta_{ij} \right]$$

$$+ (N_{ij} + 1) \log \alpha_{ij} - \frac{\hat{b}_q(\alpha_{ij}, \mathbf{x}_{-m_{ij}}) + \tilde{M}_{ij}(\alpha_{ij})}{\theta_{ij}^{\alpha_{ij}}}\ .$$

All terms of this expression are log-concave. In particular $\pi_{3;i,j}$ is log-concave by construction and the last term is a sum of concave functions of kind $g(z) \mapsto -z^{\alpha_{ij}}$. Hence the log-concavity follows from the property that the sum of concave functions is concave. Finally (4) implies the following:

$$P(J_{\tau+1} = j|\ \mathbf{j}, \mathbf{x}, u_T, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = \frac{P(J_{\tau+1} = j,\ X_{\tau+1} > u_T|\ \mathbf{j}, \mathbf{x}, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta})}{\sum_{k \in E} P(J_{\tau+1} = k,\ X_{\tau+1} > u_T|\ \mathbf{j}, \mathbf{x}, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta})}$$

$$= \frac{p_{j_\tau j} \exp\left\{ -\left(\frac{u_T}{\theta_{j_\tau j}}\right)^{\alpha_{j_\tau j}} \right\}}{\sum_{k \in E} p_{j_\tau k} \exp\left\{ -\left(\frac{u_T}{\theta_{j_\tau k}}\right)^{\alpha_{j_\tau k}} \right\}}\ .$$

The same type of argument can be used for proving (e). $\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

A.2. **JAGS implementation.** Proposition A.1 implies that JAGS should be able to run an exact Gibbs sampler. The model description we have adopted in the JAGS language based on the full likelihood (15). It is not important that the model description matches the actual model which generated the data as long as the full conditional distributions remain unchanged.

In particular, for any i, j_τ and $\sum_k N_{ik}$ can be considered as fixed. Then, given $\mathbf{p}_i$, $(N_{i1}, \ldots, N_{is})$ is regarded as a sample from a multinomial distribution with probability vector $\mathbf{p}_i$ and $\sum_k N_{ik}$ trials; furthermore, the factor $p_{j_\tau j_{\tau+1}}$ in the full likelihood is regarded as the prior for $J_{\tau+1}$ as it ranges from 1 to s; finally, upon multiplication by the Dirichlet prior for $\mathbf{p}_i$ we re-obtain the correct full conditional distribution of Proposition A.1 (a).

The uncensored inter-occurrence times are described as Weibull distributed with shape and rate parameters that depend on $\mathbf{j}$. The right-censored observation u_T is handled by a special instruction provided by the JAGS language, which should represent the actual likelihood term correctly, so that the correct form of the full conditional of $J_{\tau+1}$ is preserved.

The description of the prior distributions is routine. We have mentioned the Dirichlet prior for $\mathbf{p}_i$; α_{ij} has a shifted Gamma distribution if $m_{ij} > 1$, as specified at the end of Section 4.3; $a_{ij} := 1/\theta_{ij}^{\alpha_{ij}}$ is Gamma distributed, conditionally on α_{ij}, for any value of m_{ij} and $t_{q_{ij}}^{ij}$; $\tilde{t}_{q_{ij}}^{ij}$ is uniformly distributed if $m_{ij} = 0$, otherwise it is constant. The only nontrivial case arises for the prior of α_{ij} if $m_{ij} \leq 1$, which reduces to $\tilde{\pi}_3(\alpha_{ij}) \propto (1 - \alpha_0/\alpha)\, \mathbb{1}_{\alpha_0 \leq \alpha \leq \alpha_1}$ (see Section 4.4). This distribution is coded using the so-called zeros trick: a fictitious zero observation from a Poisson distribution with mean $\phi = -\ln\tilde{\pi}(\alpha_{ij})$ is introduced; a uniform prior over $[\alpha_0, \alpha_1]$ is assigned to α_{ij}; then the likelihood factor corresponding to the zero observation is $\exp(-\phi)$, which multiplied by the uniform prior gives the desired factor in the joint distribution of the data and the parameters.

References

Alvarez, E.E. (2005). Estimation in stationary Markov Renewal Processes, with application to earthquake forecasting in Turkey. *Methodol. Comput. Appl. Probab.*, Vol. **7**, pp. 119–130.

Berger, J.O. and Sun, D. (1993). Bayesian Analysis for the Poly-Weibull Distribution. *J. Amer. Statist. Assoc.*, Vol. **88**, pp. 1412–1418.

Betrò, B., Garavaglia, E., Grandori Guagenti, E., Rotondi, R. and Tagliani, A. (1989). Sulla distribuzione dei tempi di intercorrenza fra eventi sismici in alcune zone italiane (On the distribution of the inter-occurence times between seismic events in some Italian areas). Proceedings del 4 Conv. Naz. "L'Ingegneria Sismica in Italia", ANIDIS - Politecnico di Milano, Milano, Ed. Patron Bologna, Vol. **1**, pp. 135–144.

Bousquet, N (2006). A Bayesian analysis of industrial lifetime data with Weibull distributions. *Rapport de recherche INRIA n.6025*, pp. 1-21.

Bousquet N. (2010) Elicitation of Weibull priors. `http://arxiv.org/abs/1007.4740v2` .

Cano J., Moguerza J.M., Ríos Insua D. (2011). Bayesian analysis for semi-Markov processes with applications to reliability and maintenance. *Technical report*, Madrid, Universidad Rey Juan Carlos.

CPTI Working Group (2004). Catalogo Parametrico dei Terremoti Italiani, version 2004 (CPTI04), Ist. Naz. di Geofis. e Vulcanol., Bologna, Italy. `http://emidius.mi.ingv.it/CPTI04/`.

DISS Working Group (2007) Database of Individual Seismogenic Sources (DISS), version 3.0.2: A compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas. `http://diss.rm.ingv.it/diss/`.

Epifani I., Fortini S., and Ladelli L. (2002). A characterization for mixtures of Semi Markov processes. *Statist. Probab. Lett.*, Vol. **60**, pp. 445–457.

Foucher, Y., Mathieu, E., Saint-Pierre, P., Durand J.F. and Daurès, J.P.(2009). A Semi-Markov Model Based on Generalized Weibull Distribution with an Illustration for HIV Disease. *Biometrical Journal*, Vol. **47**, pp. 825–833.

Garavaglia, E. and Pavani, R. (2012). About Earthquake Forecasting by Markov Renewal Processes. *Methodol. Comput. Appl. Probab.*, Vol. **13**, pp. 155–169.

Gilks, W.R., Best, N.G. and Tan, K.K.C. (1995). Adaptive rejection Metropolis sampling. *Applied Statistics*, Vol. **44**, pp. 455–472.

Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, Vol. **41**, pp. 337–348.

Grandori Guagenti, E. and Molina, C. (1986). Semi-Markov Processes in seismic risk analysis. In *Semi-Markov Models: Theory and Applications*, Ed. Janssen, J., pp. 487–503, Plenum Press, New York.

Grandori Guagenti, E., Molina, C. and Mulas G. (1988). Seismic risk analysis with predictable models. *Earthquake Eng. Struct. Dynam.*, Vol. **16**, pp. 343–359.

Hougaard, P. (1999). Multi-state models: a review. *Lifetime Data Anal.*, Vol. **5**, pp. 239–264.

Limnios, N. and Oprisan, G. (2001). *Semi-Markov Model Processes*. Birkhauser, Boston.

Masala, G. (2012). Earthquakes occurrences estimation through a parametric semi-Markov approach. *J. Appl. Statist.*, Vol. **39**, pp. 81–96.

Marín J., Plà L. and Ríos Insua D. (2005). Forecasting for Some Stochastic Process Models Related to Sow Farm Management. *J. Appl. Statist.*, Vol. **32**, pp. 797–812.

Patwardhan A.S., Kulkarni R.B. and Tocher D. (1980). A semi-Markov model for characterizing recurrence of great earthquakes. *Bull. Seismol. Soc. Am.*, Vol. **70**, pp. 323–347.

Peruggia, M. and Santner, T. (1996). Bayesian analysis of time evolution of earthquakes. *J. Amer. Statist. Assoc.*, Vol. **91**, pp. 1209–1218.

Plummer, Martyn 2010. *JAGS Version 3.1.0: Just Another Gibbs Sampler*, URL http://mcmc-jags.sourceforge.net/

Rotondi, R. (2010). Bayesian nonparametric inference for earthquake recurrence time distributions in different tectonic regimes. *J. Geophys. Res.*, Vol. **115**, B01302, DOI: 10.1029/2008JB006272.

Rovida, A., Camassi, R., Gasperini P. and Stucchi M. (Eds.) (2011). CPTI11, the 2011 version of the Parametric Catalogue of Italian Earthquakes. Milano, Bologna, Italy. `http://emidius.mi.ingv.it/CPTI`.

Votsi, I. Limnios, N., Tsaklidis, G. and Papadimitriou, E. (2012). Estimation of the Expected Number of Earthquake Occurrences Based on Semi-Markov Models. *Methodol. Comput. Appl. Probab.*, Vol. **14**, pp. 685–703.

E-mail address: `ilenia.epifani@polimi.it`

POLITECNICO DI MILANO, PIAZZA LEONARDO DA VINCI 32, I-20133, MILANO, ITALY

E-mail address: `lucia.ladelli@polimi.it`

POLITECNICO DI MILANO, PIAZZA LEONARDO DA VINCI 32, I-20133, MILANO, ITALY

E-mail address: `antonio.pievatolo@mi.imati.cnr.it`

CNR IMATI, VIA BASSINI 15, I-20133 MILANO, ITALY